QUESTIONNAIRE DESIGN EFFECTS ON INTERVIEW OUTCOMES

Jeffrey Moore, Laureen Moyer, U.S. Bureau of the Census
Jeffrey Moore, U.S. Census Bureau, Center for Survey Methods Research/SRD, Washington, DC 20233

Key Words: automated instrument design, household surveys, nonresponse, interviewer assessment, respondent assessment

1. OVERVIEW

The CATI nonresponse followup phase of the Census Bureau's 1997 test panel of the American Community Survey (ACS) included an experimental test of two instrument structures — the traditional "person-based" approach versus a new "topic-based" design. The former in essence completes an entire, separate interview for each eligible household member in turn; in contrast, a topic-based interview gathers data on one "topic" for every person and then proceeds to the next topic, in effect making only one "pass" through the instrument.

Section 2 of this paper offers some basic background information about the research, including descriptions of the ACS, the characteristics of person-based and topic-based instrument designs, and the results of a small-scale pilot test of topic-based interviewing. In section 3 we summarize the design of the main experiment. Section 4 is devoted to the results of the experiment, focusing on four outcome domains: the interview process, interviewer and respondent assessments, and survey data effects. In section 5 we offer our summary and conclusions.

2. BACKGROUND

2.1 The American Community Survey

The American Community Survey (ACS) is the household survey cornerstone of the Census Bureau's new "continuous measurement" (CM) alternative to decennial census long form data collection. The primary goal of CM is to provide annual updates of detailed population and housing data throughout each decade.

Following several years of testing and development, the ACS will be implemented throughout the U.S. in 2003. When fully operational, up to three million addresses will fall into the ACS sample each year, thus enabling precise annual estimates of the housing, social, and economic characteristics for all states, as well as for all cities, counties, metropolitan areas, and population groups of 65,000 persons or more. (Smaller areas will require multiple years of data collection to reach equivalent levels of estimate precision.)

The ACS is conducted using three modes of data collection: self-enumeration through mail-out/mail-back methods; computer assisted telephone interviewing (CATI) for mail nonresponse cases for which a telephone number can be obtained; and computer assisted personal interviewing (CAPI) for a sample of mail nonresponding cases which cannot be completed by CATI. The research project described in this report evaluates an alternative design for the CATI instrument used in the middle stage of data collection.

The Census Bureau's CM website —
www.census.gov/cms/www
— contains more detailed information about the ACS and about the CM program in general.

2.2 Person-Based and Topic-Based Interviews

In its initial incarnations, the ACS/CATI survey instrument has followed a conventional person-based design for household survey questionnaires which are intended to gather data about all members of target households from a single household respondent. This design "decision" was in fact less a conscious decision than simply a direct translation of traditional paper-and-pencil questionnaire methods to the computer-assisted interview environment. The person-based approach in essence completes all topics for one person before proceeding to the next person, e.g.:

What is [person1's] sex?
What is [person1's] birth date?
What is [person1's] marital status?
Does [person1] have a work disability?
[etc. for additional topics]

What is [person2's] sex?
What is [person2's] birth date?
What is [person2's] marital status?
[etc.]

[etc. for additional persons]

The advent of computer-assisted interviewing, however, has made a "topic-based" interview sequence a viable design option. A topic-based interview completes one topic for all persons before proceeding to the next topic, e.g.:

What is [person1's] sex?
What is [person2's] sex?

[etc. for persons 3, 4,...]

What is [person1's] birth date?
What is [person2's] birth date?
[etc. for persons 3, 4, ...]

What is [person1's] marital status?
What is [person2's] marital status?
[etc. for persons 3, 4, ...]

[etc. for additional topics]

Moore (1996) summarizes the potential benefits (and a few potential pitfalls) of the topic-based approach. More recently, Couper et al. (1997) have demonstrated positive effects of the general approach. One of the key practical benefits of the topic-based interview is the freedom it allows to severely truncate the text needed for subsequent question readings after the first, full presentation of the question, e.g.:

Has [person1] ever served in the Armed Forces?
How about [person2]...?
And [person3]...?
[etc.]

2.3 Pilot Test Results
Moore and Moyer (1998) describe the results of a small-scale laboratory test of the proposed topic-based ACS CATI instrument. The primary goals of this test were to obtain insights into respondents' reactions to topic-based and person-based interviews, to refine the final topic-based design, and to develop appropriate evaluation methods for use in the large-scale field experiment.

The pilot test used paper-and-pencil prototypes of two different topic-based instruments, as well as a person-based (P) control treatment. The two topic-based alternatives varied with regard to what was considered to be a "topic." One version (T1) often grouped multiple questions (e.g., Spanish origin, race, and their accompanying detail questions) into a single topic, whereas the other more extreme design (T2) limited almost all topics to individual items.

We conducted 37 one-hour-plus laboratory sessions with respondents who met certain criteria — essentially, adult "head of household" members of large (preferably 4 or more person) households. Each paid subject responded to an ACS telephone interview (using random assignment to one of the three interview treatments, P, T1, or T2) and then answered debriefing questions about the interview. While one researcher conducted the

telephone interview, another observed the respondent's behavior through a one-way mirror, noting particularly any overt evidence of dismay, fatigue, boredom, disengagement, etc.

Overall, the pilot test results clearly suggested a general superiority of the topic-based design, and in particular the T2 topic-based design:
• *Respondent debriefing assessments:* On several dimensions pilot test respondents exhibited substantial differences in their assessments of the interview experience across the instrument treatments. For example, the T2 instrument was seen by its respondents as the least repetitive, the least likely to induce feelings of impatience, and elicited the lowest expressed preference for a different type of interview structure.
• *Respondent behavior observations:* We observed respondents' behavior through a one-way mirror during the pilot test interviews, noting any displays of pleasure or displeasure. Displays of pleasure proved to be vanishingly rare. Displeasure displays, however, were fairly common, and revealed clear differences between the instrument treatments. In particular, T2 respondents exhibited less evidence of confusion, less annoyance with and desire to speed up the interview, and less boredom/fatigue — in fact, "boredom/fatigue" displays were about 40 times more common in the person-based and T1 topic-based interviews than they were in T2 topic-based interviews.
• *Length of interview:* The pilot test did not yield strong confirmation of the expected savings in interview length with the topic-based format. Overall, the average length of the more extreme (T2) topic-based interview was 27.2 minutes per household — about 5 to 10 minutes shorter on average than the other interview treatments. However, controlling for household size reduces this difference substantially. The T2 interview averaged 10.3 minutes per interviewed adult, only about 1 to 2 minutes less than the other interview treatments. Not surprisingly, pilot test interviewers found the topic-based instruments very difficult to manage in their paper-and-pencil format; therefore, the test results were assumed to represent the minimal gains likely to be experienced with a fully automated topic-based design.

The pilot test results were judged sufficiently positive to justify a large-scale and more rigorous test of the topic-based approach, to be described in the remainder of this paper. This test used as an experimental instrument a slightly modified T2-type topic-based design that almost exclusively equated "topic" with "individual question."

3. RESEARCH DESIGN

The person-based/topic-based experiment was conducted at the Census Bureau's Jeffersonville Telephone Center (JTC) CATI facility. Interviewing was conducted from October 3 through 26, 1997 (for sample cases which had failed to respond by mail in September), and November 1 through 26, 1997 (for October mail non-respondents). Cases were assigned at random to one or the other instrument type. The two months of interviewing produced 1112 completed person-based (P) interviews and 1154 topic-based (T) interviews.

Staffing arrangements were designed to ensure to the maximum extent possible that the test of the two instruments was uncontaminated by differences in the skills or experience levels of the interviewers. Most of the interviewers (of which there were approximately 40) had general, but not ACS, experience (to avoid as much as possible any pre-set notions about the "proper" way to conduct the ACS interview), and most worked both months of the experiment (so that the interviewer pool stayed relatively constant throughout the course of the study). Interviewers were split into two teams of approximately equivalent skill levels and experience. Each team was assigned at random to either the person-based or the topic-based instrument for the first month of interviewing, and then switched to the other instrument for the second interview month.

4. RESULTS

4.1 Interview outcomes

• *Household response/refusal rates:* The experiment produced a small but statistically significant difference, favoring the topic-based instrument, in the completion of interviews among assigned, eligible cases (T=60.5%; P=56.5%), primarily due to a reduction in the proportion of cases not completed due to overt refusal to participate (T=13.0%; P=15.9%). (Both differences are significant according to simple t-tests — t=2.53 and 2.56, respectively; p<.05.)

The significant refusal rate advantage enjoyed by the topic-based instrument presents an interesting logical puzzle. As is typically the case (Groves and Couper, 1998), mid-interview "breakoffs" in this study were very rare; virtually all refusals occurred during pre-interview "negotiations," well before the structure of the interview was even potentially apparent to the refusers. Thus, the refusal rate difference between the instrument treatments in this test seems most likely to have arisen from differences in interviewers' behaviors in the face of similar base rates of respondent reluctance to participate.

• *Interview length:* Across all completed cases, topic-based interviews averaged 32.6 minutes in length, about 2 minutes less than person-based interviews. This difference is both statistically significant (t=3.99, p<.001) and, given the eventual scale of the ACS operation, economically significant as well. Moore (1996) estimates that reducing the average CATI interview length by this amount in the full-production ACS environment would yield annual savings of approximately $300,000 in interviewer labor costs alone.

4.2 Interviewers' assessments

Interviewers completed debriefing questionnaires approximately two weeks into each interview month. The two administrations enabled us to obtain both the "naive" opinions of the interviewers about each instrument, before they had any experience with the other instrument, and their opinions at the end of the experiment, when they could directly compare the alternate designs. Here we summarize the responses to the debriefings. (Note that, because of the nature of the data and the small number of cases, we treat these results more as impressionistic observations than as statistical data to be analyzed with statistical tests.)

• *Interviewers' likes and dislikes:* The debriefing questionnaires included open-ended questions asking interviewers what they liked about the instrument assigned to them that month and what they disliked about that instrument. Interviewers had many more positive things to say about the topic-based instrument than the person-based instrument, and many more negative things to say about the person-based instrument than the topic-based instrument. Virtually all interviewers offered at least one "like" comment about the topic-based instrument, compared to only about half for the person-based instrument, and those who offered "like" comments to the topic-based instrument provided about 50% more such comments on average (1.8) than did those who provided "like" comments to the person-based instrument (1.2). The results regarding "dislike" comments present an almost perfect mirror image to the "like" results.

The main reason offered for liking the topic-based instrument was that it made for a faster interview, and one that reduced interviewers' effort substantially. The most common reasons offered for liking the person-based instrument had to do with its greater structure and "orderliness." For both instruments, however, the foci of the "dislike" responses were far more concentrated. The person-based instrument was seen as too long, and highly repetitive and tedious; the topic-based instrument was faulted for not working as well in roommate/boarder households and other situations requiring callbacks to

complete the interview.

• *Interviewers' ratings of design advantages:* The debriefing questionnaires asked interviewers to rate on five different dimensions the extent to which their assigned instrument contributed to an improved interview. Each of these items used a 5-point rating scale ranging from "strongly agree" to "strongly disagree." Table 1 shows the proportion of positive ("strongly agree" plus "agree") responses for each of the five dimensions. By generally overwhelming margins, the interviewers favored the topic-based instrument as the one that made the interview flow more quickly and naturally, helped them become familiar with the organization of the questionnaire, made it easier for them to conduct the interview, and made it easier for them to probe for more correct answers.

• *Interviewers' ratings of instrument performance in different types of households:* Interviewers also rated the performance of their assigned instrument design in various types of households. Again, we used 5-point scales ranging from "very well" to "very poorly." Table 2 shows the proportion of interviewers who rated their instrument positively (i.e., who gave a response of "very well" or "well") for each type of household. These results clearly indicate interviewers' general preference for the topic-based instrument in all types of households (with the exception of single-person households, where the two instruments actually functioned identically). The results for households with unrelated persons and elderly households are especially interesting. In both cases, contrary to the comments interviewers made during a focus group (see Moore and Moyer, 1998) and in their open-ended "like/dislike" debriefing replies (see above), interviewers' debriefing questionnaire responses suggest that they perceived the topic-based instrument to have performed better than the person-based version.

4.3 Respondents' assessments

Most telephone interviews included a set of "respondent debriefing" questions, administered after the completion of the main interview, the purpose of which was to assess respondents' reactions to the interview. Cases were excluded from the debriefing for the following reasons: (1) no permanent residents at the address (this was the case in about 1% (n=28) of the 2266 completed interviews); (2) the household contained only one resident (502 interviews were excluded from the debriefing for this reason); and (3) the ACS interview was interrupted by a call-back or a switch of respondents (this condition excluded another 236 cases). Thus, the final analysis sample for the respondent debriefing data consists of 1500 cases — 714 in the person-based

treatment, and 786 in the topic-based.

Responses to the debriefing questions suggest several ways in which those interviewed with the person-based instrument and those interviewed with the topic-based instrument experienced the interview differently. In general, the results suggest a preference for the topic-based instrument. For example, compared to person-based respondents, topic-based respondents were more likely to report that they stayed interested throughout the interview, and overwhelmingly less likely, when presented with the option, to express a preference for the other instrument structure.

In some cases, the debriefing responses were found to be affected by the relatedness of household members. For example, there was no significant difference overall in respondents' tendency to label the ACS/CATI questions as "repetitious." However, this apparent similarity masked substantial differences according to the relatedness of household members. In households whose members were all related to each other (which comprised about 93% of the debriefing cases), topic-based respondents were less likely than person-based respondents to view the interview as "repetitious;" in non-related households the reverse was true.

4.4 Survey data outcomes

We also used the responses produced (or, in the case of item nonresponse, not produced) by the two ACS instruments to assess the impact of instrument design. In this section we examine two such data outcomes. The first, item nonresponse, presents the most unambiguous evidence concerning the differential effects of the person-based and topic-based instruments on data quality — in fact, in the absence of validating data, it is the only unambiguous evidence. The other comparison examines instrument differences in the tendency to produce consistent reports for all household members on some characteristic (e.g., race, language spoken at home, etc.)

• *Item nonresponse:* We examined instrument effects on item nonresponse through an item-by-item comparison of missing data rates, focusing on all items for which either instrument's nonresponse rate exceeded 2%. (In other words, we ignored any nonresponse differences where the overall level of nonresponse was trivial.) The results of this analysis clearly indicate anadvantage to the topic-based instrument. Among the 43 items with "important" levels of nonresponse, 29 show significant non-response differences by instrument treatment. For 24 of those 29 differences the non-response rate is lower for the topic-based instrument, versus only 5 differences in favor of the person-based

instrument (see Moore and Moyer, 1998, for details).

• *Within-household response consistency:* We also looked for instrument differences in the propensity to produce within-household consistency on certain individual-level characteristics. This analysis was motivated by concerns among some ACS analysts that the topic-based format might tend to gloss over differences among household members, producing inflated levels of within-household consistency. (At the same time, however, others conjectured that any differences in within-household consistency might just as well be attributed to a tendency of the person-based design to produce false differences among household members where there should be consistency.)

We analyzed within-household consistency differences by type of instrument for six characteristics. For three of the characteristics — race, Hispanic origin, and current school enrollment — there was no difference between the person-based and topic-based treatments in the rate of within-household consistency. Three other characteristics did show significant differences:

(1) Citizenship. Contrary to the primary concern of ACS subject matter analysts, the person-based format elicited consistent reporting of citizenship status in 86% of interviewed households, a significantly <u>higher</u> level of consistency than the 83% rate in households responding to the topic-based instrument;

(2) Mobility. The within-household consistency results for residency in the current living quarters five years ago show a higher level of consistency for topic-based households (90%) than for person-based households (85%); and

(3) Use of a non-English "at home" language. Again, compared to person-based interviews, topic-based interviews elicited more frequent within-household consistency on use of a language other than English at home (94% vs. 91%, respectively).

Thus, although this analysis identified some significant results, they were of inconsistent direction. Combined with the absence of differences for the other comparisons, we are forced to conclude that there is no consistent effect of instrument type on uniform reporting of various characteristics among all household members.

5. SUMMARY AND CONCLUSIONS

The results of the person/topic experiment suggest mostly quite positive outcomes for the topic-based design — a higher response rate, a lower refusal rate, reduced interview length, more favorable evaluations from interviewers and respondents, and generally lower rates of item nonresponse. The positive outcomes derived from the topic-based approach offer strong justification for a change to the topic-based design in the ACS mail nonresponse followup system. Some lingering questions remain, however, which future research will need to address.

One question concerns the meaning of observed differences in some of the response distributions (data not shown here — see Moore and Moyer, 1998), and the differences in the tendency of the two designs to elicit consistent responses from all household members on certain characteristics. The major issue here, of course, is whether these <u>response</u> differences imply data <u>quality</u> differences, and, if so, which instrument produces higher quality data. Additional assessments will be needed to better understand these phenomena and their implications, if any, for data quality differences.

A second high-priority question concerns the item nonresponse results in the field experiment. Although the vast majority of item nonresponse differences favored the topic-based design, two items for which the topic-based instrument produced significantly <u>more</u> item nonresponse were wage/salary income and total income (see Moore and Moyer, 1998, for details). Because of the importance of income data to the ACS, these results are of some concern to CM staff, even though they run counter to the overall nonresponse results. Again, additional research will be required to assess the extent to which these differences might be due to sample differences, instrument design flaws, or some other factors extrinsic to the topic-based design; or whether they are, in fact, an inherent weakness of a topic-based interview.

Finally, as noted earlier, the CATI operation is only the first stage of mail nonresponse followup in the ACS. CATI in fact produces only about half of all of the ACS interviews obtained from mail nonrespondent households; the remainder are obtained via personal-visit CAPI interviewing. Therefore, it is important to determine whether the beneficial effects of the topic-based approach observed in the CATI setting carry over to a CAPI mode of administration.

REFERENCES

Couper, M., M. Fuchs, S. Hansen, and P. Sparks (1997). "CAPI Instrument Design for the Consumer Expenditure (CE) Quarterly Interview Survey." University of Michigan, Ann Arbor, MI, final report to the Bureau of Labor Statistics (Task Order #12, Contract J-9-J-5-0025), December, 1997.

Groves, R. and M. Couper (1998). Nonresponse in Household Interview Surveys. New York: John Wiley & Sons.

Moore, J. (1996). "Person- vs. Topic-Based Design for Computer-Assisted Household Survey Instruments." Paper presented at the International Conference on Computer-Assisted Survey Information Collection, San Antonio, TX, December 11-14, 1996.

Moore, J. and L. Moyer (1998). "ACS/CATI Person-Based/Topic-Based Field Experiment." Unpublished U.S. Census Bureau report, July 29, 1998.

Table 1: Interviewers' Assessments of Instrument Design Advantages by Instrument Type		

"Please rate the [person/topic]-based instrument. Did organizing the questions so that all questions were asked
    (P) for one person before moving on to the next person ...
    (T) for all persons for a given topic before moving on to the
      next topic ...
	% "Strongly Agree" or "Agree"	

	Person-Based (n=36)	Topic-Based (n=33)
... make interviews flow quickly?"	25%	94%*
... make interviews flow naturally?"	22%	94%
... help you become familiar with the organization of the instrument?"	42%	84%*
... make it easy to conduct an interview?"	42%	94%
... make it easier to probe for more correct answers?"	47%	73%

*Note: One debriefing questionnaire lacked a response to the starred items, reducing the denominator to 32.

Table 2: Interviewers' Assessments of Instrument Performance in Various Kinds of Households by Instrument Type		
	% Reporting "Very Well" or "Well"	
"Please rate how you felt the [person/topic]-based instrument performed with different types of households."	Person-Based % (n)	Topic-Based % (n)
Single Person Households	100% (35)	73% (33)
2-3 Person Households	71% (34)	79% (34)
4+ Person Households	9% (33)	85% (34)
Households with Children	44% (32)	97% (34)
Households with Unrelated Persons	32% (28)	58% (33)
Reluctant/Unenthusiastic Respondents	3% (33)	62% (29)
Elderly Respondents	24% (33)	71% (34)